



Public Sentiment Analysis Towards Mayor Teddy Indra Wijaya with Logistic Regression Approach

Analisis Sentimen Publik Terhadap Mayor Teddy Indra Wijaya dengan Pendekatan Logistic Regression

Antonius Steven Dimas Prasetya Sinaga¹, Adam Sekti Aji²

^{1,2}Program Studi Informatika, Fakultas Sains dan Teknologi
Universitas Teknologi Yogyakarta, Indonesia

E-Mail: ¹stevensinaga26@gmail.com, ²adamaji@staff.uty.ac.id

Received Oct 4th 2024; Revised Nov 21th 2024; Accepted Dec 10th 2024; Available Online Nov 15th 2024

Corresponding Author: Antonius Steven Dimas Prasetya Sinaga

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This study aims to analyze public sentiment toward Mayor Teddy Indra Wijaya, an aide to Prabowo Subianto, using the Logistic Regression algorithm. Sentiment analysis is essential as public figures like Mayor Teddy, who are closely associated with prominent political leaders, significantly shape public perceptions. Understanding these sentiments is crucial to identifying positive and negative public responses, which serve as indicators of public perception regarding Prabowo Subianto's political image. This research utilizes data from the Twitter social media platform, collected through web crawling from January 2024 until the time of this report. The primary objective of this study is to explore the effectiveness of the Logistic Regression algorithm in mapping public sentiment based on complex and diverse textual data. The research process includes data collection, preprocessing, labeling, model training, and performance evaluation using metrics such as accuracy, precision, recall, and F1 score. The findings reveal that the Logistic Regression model achieved an accuracy of 73.68%, a precision of 75%, a recall of 71%, and an F1 score of 0.73. These results demonstrate the model's considerable effectiveness in predicting public sentiment. This study contributes to the development of sentiment analysis methodologies in socio-political contexts and provides strategic insights for managing public communication and reputation in the digital era.

Keyword: Logistic Regression, Mayor Teddy, Sentiment Analysis, Social Media, Twitter

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Mayor Teddy Indra Wijaya, ajudan Prabowo Subianto, dengan menggunakan algoritma Logistic Regression. Analisis sentimen diperlukan karena keberadaan tokoh publik seperti Mayor Teddy, yang berasosiasi erat dengan figur politik nasional, dapat memengaruhi persepsi publik secara luas. Pemahaman terhadap sentimen ini penting untuk mengidentifikasi respons positif dan negatif masyarakat, yang kemudian dapat digunakan sebagai indikator persepsi publik terhadap citra politik Prabowo Subianto. Penelitian ini memanfaatkan data dari platform media sosial Twitter yang dikumpulkan melalui crawling selama periode Januari 2024 hingga saat laporan ini disusun. Masalah yang ingin diselesaikan adalah bagaimana algoritma Logistic Regression dapat digunakan secara efektif untuk memetakan sentimen publik berdasarkan data teks yang kompleks dan beragam. Proses penelitian meliputi pengumpulan data, preprocessing, pelabelan, pelatihan model, dan evaluasi performa model menggunakan metrik seperti akurasi, precision, recall, dan F1 score. Hasil penelitian menunjukkan bahwa model Logistic Regression yang digunakan memiliki akurasi sebesar 73,68%, precision 75%, recall 71%, dan F1 score 0,73. Hasil ini menunjukkan kinerja yang cukup baik dalam memprediksi sentimen masyarakat. Penelitian ini diharapkan dapat berkontribusi pada pengembangan metode analisis sentimen dalam konteks sosial-politik serta memberikan masukan strategis bagi pengelolaan komunikasi publik dan reputasi di era digital.

Kata Kunci: Logistic Regression, Mayor Teddy, Sentiment Analysis, Social Media, Twitter

1. PENDAHULUAN

Dalam beberapa waktu terakhir, Mayor Teddy Indra Wijaya, ajudan Menteri Pertahanan Prabowo Subianto, menjadi perhatian publik. Sebagai perwira Kopassus dengan kedekatan profesional dengan Prabowo,

Mayor Teddy tidak hanya menjalankan tugas resmi, tetapi juga menjadi bagian penting dari citra politik pemimpin nasional. Kehadirannya dalam berbagai acara resmi sering kali menciptakan narasi di media sosial, yang dapat memengaruhi persepsi publik terhadap Prabowo Subianto. Fenomena ini menegaskan pentingnya analisis sentimen untuk memahami respons masyarakat terhadap tokoh pendukung dalam ranah politik nasional [1]. Dengan begitu, penelitian ini relevan untuk mendukung pengelolaan citra publik figur di era digital [2].

Urgensi penelitian ini terletak pada pentingnya memahami opini masyarakat terhadap figur yang berasosiasi langsung dengan tokoh politik utama. Media sosial, sebagai ruang ekspresi opini, memungkinkan masyarakat menyampaikan pendapat secara terbuka, baik berupa dukungan maupun kritik [3]. Analisis sentimen terhadap opini ini dapat memberikan wawasan strategis untuk mengelola citra publik figur secara lebih efektif [4]. Selain itu, penelitian ini juga relevan dalam konteks ilmu sosial-politik untuk mengevaluasi bagaimana opini masyarakat terbentuk dan memengaruhi reputasi pemimpin politik.

Permasalahan yang dihadapi adalah bagaimana memetakan sentimen publik secara akurat dari data media sosial yang kompleks dan tidak terstruktur. Data Twitter, misalnya, memiliki volume besar dengan karakteristik teks yang tidak konsisten, sehingga diperlukan metode yang mampu mengolah data ini secara efektif [5]. Logistic Regression dipilih dalam penelitian ini karena kemampuannya menangani data teks dengan distribusi sederhana dan korelasi antar fitur [6]. Algoritma ini juga dikenal efisien, memiliki interpretasi yang jelas, dan mudah diimplementasikan menggunakan pustaka pembelajaran mesin yang tersedia luas.

Pemilihan Logistic Regression didukung oleh kelebihan algoritma ini dibandingkan metode lain seperti Naïve Bayes atau Support Vector Machine (SVM) [7]. Naïve Bayes, meskipun efisien, sering kali sensitif terhadap distribusi data dan kurang fleksibel untuk data dengan korelasi antar fitur [8]. Sementara itu, SVM membutuhkan sumber daya komputasi yang lebih besar, terutama pada dataset besar, dan kurang interpretatif dibandingkan Logistic Regression [9]. Dengan demikian, Logistic Regression memberikan keseimbangan antara efisiensi, interpretabilitas, dan akurasi untuk penelitian ini [10].

Penelitian [11] ini menganalisis sentimen publik di Twitter terhadap Puan Maharani sebagai kandidat presiden menggunakan kombinasi algoritma Naïve Bayes, SMOTE, dan AdaBoost. Hasil penelitian menunjukkan kombinasi algoritma ini meningkatkan Area Under Curve (AUC) menjadi 0.733 dibandingkan dengan Naïve Bayes saja yang hanya mencapai 0.625 [12]. Namun, akurasi dan recall sedikit menurun karena kompleksitas algoritma yang lebih tinggi. Penelitian ini menunjukkan efektivitas pendekatan ensemble dalam analisis sentimen, tetapi tidak fokus pada interpretasi linguistik atau analisis tematik. Dalam penelitian ini, Logistic Regression dipilih untuk menghasilkan model yang lebih sederhana dan interpretatif, yang lebih sesuai untuk data teks dengan distribusi biner.

Mengklasifikasikan sentimen masyarakat terhadap UMKM di Twitter menggunakan algoritma Naive Bayes Classifier (NBC). Penelitian [13] ini menggunakan library Python TextBlob untuk preprocessing data dan menghitung hasil secara manual dengan NBC, yang memberikan hasil akurasi tinggi. Meskipun sederhana dan efisien, penelitian ini terbatas pada penggunaan algoritma tunggal tanpa perbandingan dengan metode lainnya. Dalam penelitian ini, Logistic Regression dipilih karena memberikan fleksibilitas lebih baik dibandingkan Naïve Bayes, terutama dalam menangani data dengan korelasi antar fitur, sehingga dapat meningkatkan kemampuan model untuk memprediksi sentimen publik [14].

Penelitian [15] ini membandingkan berbagai metode analisis sentimen, termasuk Support Vector Machine (SVM), Lexicon, Naïve Bayes, Random Forest, Decision Tree, dan K-Nearest Neighbor (KNN), menggunakan data ulasan aplikasi KAI Access dari Google Play Store [16]. Confusion matrix digunakan untuk visualisasi hasil analisis dan evaluasi akurasi. Penelitian ini memberikan wawasan mendalam tentang kinerja berbagai algoritma pada dataset ulasan aplikasi, tetapi fokusnya lebih pada perbandingan model daripada optimasi algoritma untuk analisis sentimen spesifik. Penelitian ini berbeda karena fokus pada Logistic Regression sebagai metode tunggal untuk mendalami analisis persepsi masyarakat terhadap tokoh politik, menggunakan data sosial yang bersifat lebih kompleks daripada ulasan aplikasi.

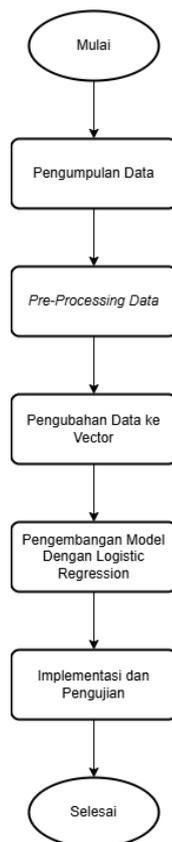
Penelitian [17] menggunakan Naïve Bayes untuk menganalisis sentimen review produk Shopee, dengan visualisasi tambahan berupa WordCloud. Akurasi tertinggi (85%) dicapai pada data training sebesar 60%, dengan penggunaan fitur pembobotan TF-IDF. Penelitian ini menunjukkan efektivitas Naïve Bayes pada data e-commerce yang terstruktur, tetapi kurang cocok untuk data sosial yang lebih beragam. Penelitian ini berbeda karena menggunakan Logistic Regression yang lebih fleksibel dalam menangani pola linguistik dari data media sosial yang kompleks, memungkinkan analisis lebih mendalam terhadap opini publik terhadap figur politik [18].

Dibandingkan dengan penelitian terdahulu, penelitian ini memiliki beberapa perbedaan signifikan dan kelebihan utama. Penelitian sebelumnya seperti [11] dan [17] lebih fokus pada analisis sentimen di sektor non-politik, seperti e-commerce dan opini terhadap produk tertentu, yang datanya relatif lebih terstruktur. Sebaliknya, penelitian ini mengkaji persepsi publik terhadap figur politik, yaitu Mayor Teddy Indra Wijaya, yang berasosiasi langsung dengan citra politik Prabowo Subianto, menggunakan data media sosial yang kompleks dan tidak terstruktur. Sementara metode seperti Naïve Bayes banyak digunakan dalam penelitian sebelumnya, Logistic Regression dipilih dalam penelitian ini karena kemampuannya menangani korelasi antar

fitur dan menghasilkan model yang lebih interpretatif [19]. Penelitian ini juga memberikan kontribusi baru dengan fokus pada konteks sosial-politik Indonesia, menawarkan wawasan strategis bagi pengelolaan citra publik figur dalam ranah politik, yang belum dibahas secara mendalam dalam penelitian terdahulu [20].

2. MATERIALS AND METHOD

Penelitian ini menggunakan metode analisis sentimen dengan model Logistic Regression, yang merupakan salah satu algoritma populer dalam pembelajaran mesin untuk klasifikasi biner [21]. Seiring dengan meningkatnya penggunaan media sosial, analisis sentimen telah menjadi sangat penting, karena platform digital menjadi ruang utama bagi masyarakat untuk mengekspresikan opini mereka dalam bentuk teks [22]. Dalam konteks ini, Logistic Regression dipilih sebagai pendekatan untuk menganalisis data teks yang besar dan tidak terstruktur, mengklasifikasikan sentimen menjadi kategori positif atau negatif [23]. Pendekatan ini memungkinkan pemahaman yang lebih mendalam mengenai persepsi publik terhadap figur politik atau peristiwa tertentu. Proses implementasi analisis sentimen menggunakan Logistic Regression melalui beberapa tahapan utama, seperti yang terlihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

2.1 Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data dari berbagai platform media sosial yang relevan, terutama Twitter, untuk mendapatkan pandangan umum dari pengguna mengenai Mayor Teddy. Data ini dikumpulkan melalui teknik *web scraping* atau crawling dengan menggunakan kata kunci tertentu yang relevan dengan penelitian ini. Pengumpulan data ini menghasilkan kumpulan teks yang berisi opini, komentar, atau tanggapan publik terkait Mayor Teddy. Proses ini bertujuan untuk memperoleh dataset yang cukup besar dan representatif untuk analisis selanjutnya.

2.2 Preprocessing Data

Setelah data terkumpul, tahap berikutnya adalah preprocessing data untuk memastikan kualitas dan konsistensi data yang akan digunakan dalam analisis. Tahap preprocessing ini sangat penting untuk menghilangkan gangguan dalam data dan mempersiapkan teks untuk analisis yang lebih akurat. Proses preprocessing ini meliputi beberapa langkah penting:

1. Melakukan pembersihan teks, yaitu menghapus karakter-karakter yang tidak relevan seperti tanda baca, angka, dan simbol yang tidak penting.

2. Melakukan tokenisasi yaitu memecah teks menjadi kata atau token yang lebih kecil agar model dapat memprosesnya.
3. Melakukan penghapusan Stopwords yaitu menghapus kata-kata yang tidak memiliki makna penting dalam analisis sentimen, seperti "dan", "atau", "di", dan sebagainya.
4. Melakukan lematisasi yaitu mengubah kata-kata menjadi bentuk dasarnya untuk mengurangi variasi kata yang berarti sama, misalnya mengubah "berlari" menjadi "lari".

2.3 Pengubahan Data ke Vektor

Setelah data dibersihkan, tahap selanjutnya adalah mengubah teks menjadi representasi numerik dalam bentuk vektor. Hal ini dilakukan dengan menggunakan teknik *vectorization*, seperti *Term Frequency-Inverse Document Frequency* (TF-IDF), yang mengukur pentingnya suatu kata dalam dokumen berdasarkan frekuensinya dan kemunculannya dalam seluruh korpus. Dengan mengubah teks menjadi bentuk numerik, model dapat memproses data tersebut dan menganalisis pola sentimen yang terkandung di dalamnya. Proses *vectorization* ini memungkinkan model untuk memahami hubungan antar kata dan pola linguistik dalam data.

2.4 Pengembangan Model dengan Logistic Regression

Setelah data diubah menjadi vektor, tahap berikutnya adalah pengembangan model menggunakan Logistic Regression. Model ini dilatih dengan data latih yang telah diproses, untuk mengidentifikasi pola-pola sentimen berdasarkan fitur yang ada dalam data. Logistic Regression adalah metode yang cocok untuk klasifikasi biner, dalam hal ini untuk mengklasifikasikan sentimen menjadi positif atau negatif. Pelatihan model dilakukan dengan mengoptimalkan parameter model menggunakan algoritma optimisasi, seperti *Gradient Descent*, untuk meminimalkan fungsi kerugian (*loss function*).

2.5 Implementasi dan Pengujian Model

Setelah model dilatih, tahap selanjutnya adalah pengujian untuk mengevaluasi akurasi dan kinerja model. Data uji, yang belum pernah dilihat oleh model selama pelatihan, digunakan untuk mengukur seberapa baik model dapat memprediksi sentimen dengan data baru. Pengujian ini juga mencakup penghitungan metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk menilai kualitas model. Model yang telah diuji kemudian diimplementasikan dalam lingkungan produksi untuk menganalisis data baru secara berkelanjutan. Model ini akan digunakan untuk memberikan wawasan tentang opini publik yang terus berkembang terkait Mayor Teddy dan pengaruhnya terhadap citra Prabowo Subianto.

2.6 Evaluasi dan Penerapan Model

Tahap akhir adalah penerapan model dalam analisis data sentimen secara lebih luas. Setelah pengujian dan evaluasi, model yang sudah terlatih dan teruji digunakan untuk menganalisis sentimen masyarakat yang lebih besar dan lebih beragam, membantu dalam pengambilan keputusan berbasis data sentimen. Implementasi ini juga akan memberikan masukan bagi strategi komunikasi politik yang lebih efektif dalam mengelola citra tokoh politik di media sosial.

3. HASIL DAN PEMBAHASAN

3.1. Crawling Data

Penelitian ini dilaksanakan dengan mengumpulkan data komentar dari platform media sosial X (dulu dikenal sebagai Twitter) sebanyak 218 data. Data yang terkumpul kemudian akan dibersihkan dengan menghapus elemen-elemen yang tidak relevan, seperti spasi kosong, mention pengguna, dan bagian lain yang tidak diperlukan. Setelah proses pembersihan, data tersebut disimpan dalam format CSV. Pengumpulan data dilakukan dengan menggunakan kata kunci "Mayor Teddy" untuk mendapatkan konteks yang sesuai dengan penelitian ini, seperti yang terlihat pada Tabel 1.

Tabel 1. Contoh Hasil Crawling Data Twitter

NO	TWEET
1	Momen duet Menhan Prabowo Subianto dan Mayor Teddy dalam acara halalbihalal bersama Akabri jadi sorotan bikin gemes emak-emak! #GarudaTV #Laporan8 #MenhanPrabowo #MayorTeddy https://t.co/LPplPfl2Fu
2	Nikita Mirzani secara terbuka membeberkan perilaku ajudan Prabowo Subianto seperti termasuk Mayor Teddy hingga mantan kekasihnya Rizky Irmansyah. #NikitaMirzani #MayorTeddy #PrabowoSubianto #MedcomId https://t.co/2jnTRB8r7a

Format CSV digunakan untuk menyimpan dan membuka data yang akan diproses menggunakan Python dengan bantuan pustaka Pandas DataFrame. Pemilihan format CSV didasarkan pada kemampuannya untuk menyimpan data dalam bentuk tabel yang efisien dan sederhana, di mana nilai-nilai dipisahkan oleh tanda koma. Hal ini memudahkan pengguna dalam mengelola dan memahami struktur data.

3.2. Data Preprocessing

Proses preprocessing bertujuan untuk mencegah kehilangan informasi, mengurangi redundansi, dan mengatasi ketidakkonsistenan dalam data yang diperoleh melalui proses scraping. Data yang diambil sering kali mengandung ketidakkonsistenan, seperti variasi dalam gaya penulisan, penggunaan singkatan, dan simbol yang tidak relevan. Oleh karena itu, langkah-langkah preprocessing menjadi penting untuk memastikan bahwa teks menjadi lebih seragam dan mudah dianalisis, seperti yang terlihat pada Tabel 2.

Tabel 2. Contoh Data Hasil Scraping yang Akan Dilakukan Preprocessing

NO	Komentar
1	Jamaah Perempuan Ini Bawa Foto Mayor Teddy di Depan Kabah #suaradotcom #Haji2024 #JamaahHaji #MayorTeddy #viral https://t.co/ecEjp75gqj Senior dan Junior dalam satu frame @SudaryonoSudar @teddsky_89 #jatengmajumapan #masdar
2	#sudaryono #jateng1 #calongubernur #darihati #darlink #satukomando #jatengmaju #dar #dengar #amati #reaksi #dpdgerindrajateng #mayorteddy #mayorteddyindrawijaya #gerindrakotamagelang https://t.co/TGhlsAajqZ

Hasil dari proses *scraping* menunjukkan beragam komentar yang mencerminkan pandangan publik terhadap Mayor Teddy. Misalnya, beberapa komentar menyoroti kehadiran Mayor Teddy dalam berbagai acara, sementara yang lain mencerminkan interaksi sosial dan pernyataan pribadi mengenai sosoknya. Komentar-komentar ini mengandung elemen-elemen seperti tagar, mention, dan tautan yang perlu dibersihkan agar analisis selanjutnya lebih akurat. Data yang telah diambil meliputi komentar-komentar yang dapat diproses lebih lanjut untuk mendapatkan wawasan yang lebih mendalam mengenai sentimen publik.

3.2.1. Text Cleaning

Langkah pertama dalam tahap *preprocessing* penelitian ini adalah melakukan *text cleaning*, yang bertujuan untuk menghapus elemen-elemen yang tidak relevan atau tidak diinginkan dari teks mentah. Proses ini meliputi penghilangan karakter khusus, tanda baca yang berlebihan, angka, serta spasi kosong yang tidak perlu. Setelah tahap ini, teks menjadi lebih bersih dan siap untuk langkah-langkah berikutnya, sehingga memudahkan dalam analisis lebih lanjut, seperti yang terlihat pada Tabel 3.

Tabel 3. Contoh Komentar Setelah Dilakukan Text Cleaning

Komentar Sebelum	Komentar sesudah text cleaning
Jamaah Perempuan Ini Bawa Foto Mayor Teddy di Depan Kabah #suaradotcom #Haji2024 #JamaahHaji #MayorTeddy #viral https://t.co/ecEjp75gqj Senior dan Junior dalam satu frame @SudaryonoSudar @teddsky_89 #jatengmajumapan #masdar #sudaryono #jateng1 #calongubernur #darihati #darlink #satukomando #jatengmaju #dar #dengar #amati #reaksi #dpdgerindrajateng #mayorteddy #mayorteddyindrawijaya #gerindrakotamagelang https://t.co/TGhlsAajqZ	jamaah perempuan ini bawa foto mayor tedy di depan kabah senior dan junior dalam satu frame

Hasil akhir dari *text cleaning* menunjukkan bahwa teks yang sebelumnya berantakan kini telah menjadi lebih sederhana dan fokus pada informasi utama. Misalnya, komentar yang semula panjang dan mengandung banyak informasi tambahan kini diubah menjadi kalimat yang lebih langsung, seperti menghilangkan hashtag dan tautan yang tidak perlu.

3.2.2. Tokenizing

Tahap kedua dalam preprocessing adalah tokenizing, yang merupakan proses memecah teks atau dokumen menjadi unit-unit terkecil yang disebut "token." Token ini bisa berupa kata, frasa, atau karakter, bergantung pada tingkat detail yang ingin dicapai. Melalui proses ini, setiap teks yang telah dibersihkan dipecah menjadi kata-kata yang terpisah, memungkinkan analisis kata secara individual, seperti yang terlihat pada Tabel 4.

Tabel 4. Hasil Implementasi Tokenizing

Komentar Sebelum	Komentar Sesudah
jamaah perempuan ini bawa foto mayor tedy di depan kabah senior dan junior dalam satu frame beginilah polah unik mayor tedy kalau diajak foto bareng abangnya maunya dia yang dirangkul gimana fans candu ya suaranya kalian pengen request mayor tedy ngomong apa tulis di reply ya dar	jamaah.perempuan,ini,bawa,foto,mayor,tedy,di,depan,kabah senior,dan,junior,dalam,satu,frame beginilah,polah,unik,mayor,tedy,kalau,diajak,foto,bareng ,abangnya,maunya,dia,yang,dirangkul,gimana,fans,candu,ya ,suaranya,kalian,pengen,request,mayor,tedy, ngomong,apa,tulis,di,reply,ya,dar

Setelah langkah *tokenizing*, setiap komentar kini terpisah menjadi kata-kata individu. Proses ini memungkinkan untuk mengidentifikasi dan menganalisis frekuensi serta pola penggunaan kata dalam komentar, yang akan sangat membantu dalam analisis sentimen di tahap selanjutnya

3.2.3. Filtering

Tahap selanjutnya adalah *stopword removal*, yang merupakan proses penyaringan kata-kata dalam dokumen dengan tujuan mengurangi jumlah kata dalam korpus dengan menghapus kata-kata yang tergolong stopwords. Stopwords adalah kumpulan kata yang cenderung kurang informatif dan sering muncul dalam dokumen, seperti kata ganti, kata penghubung, dan kata-kata yang tidak memiliki makna penting untuk menentukan topik dokumen. Stopwords ini biasanya terdiri dari kata-kata seperti kata ganti orang, kata penghubung, kata seruan, kata pertanyaan, dan kata-kata lainnya yang tidak memiliki makna penting dalam penentuan topik dokumen.

Setelah melakukan *stopword removal*, jumlah kata yang tersisa menjadi lebih sedikit dan lebih fokus pada informasi yang penting, yang mendukung analisis sentimen dengan menghilangkan kata-kata yang tidak memberikan kontribusi signifikan terhadap pemahaman makna teks. Dengan mengurangi kata-kata yang tidak perlu, analisis menjadi lebih efisien dan relevan. Proses ini dilakukan dengan memanfaatkan *library* Python yang dikenal sebagai Sastrawi.

3.3. Labelling

Labelling dalam Natural Language Processing (NLP) adalah langkah kritis yang melibatkan penambahan informasi klasifikasi pada data teks untuk berbagai tujuan analisis. Klasifikasi teks adalah salah satu aplikasi utama dari proses pelabelan, di mana teks diberi label untuk membedakan kategori atau kelas tertentu, seperti dalam pengenalan spam email atau kategorisasi artikel berita, pada tabel 5 merupakan hasil dari *Labelling* nya.

Tabel 5. Hasil *Labelling*

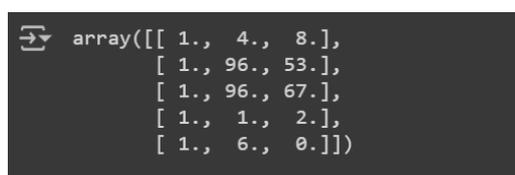
No	Full Text	Labels	Sentiment
1	[ahir, kaki, tangan, mata, baris, gigi, hilang, alami, marah, belakang]	-1	Negative
2	[damping, prabowo, kampanye, staf, mku, 2024, mk, putus, ri4, mayor, tddy, melanggar]	-1	Negative
3	[mk, hadir, mayor, tddy, debat, capres, langgar, atur, mku]	1	Positive
4	[olang, putus, mk, aries, baswedan, spekulasi]	-1	Negative
5	[hakim, mk, hadir, mayor, tddy, debat, capres, nilai, netral, wii, ni]	0	Positive

Tabel 6 menunjukkan distribusi label sentimen pada tiap cuitan yang dianalisis dalam penelitian ini. Sentimen tersebut dikategorikan menjadi tiga kelompok utama: positif, netral, dan negatif. Dari total cuitan yang dianalisis, terdapat 34 cuitan dengan sentimen positif, 102 cuitan dengan sentimen netral, dan 52 cuitan dengan sentimen negatif.

3.4. Feature Extractions

Feature extraction merupakan langkah penting dalam analisis data yang bertujuan untuk mengidentifikasi dan mengekstrak fitur-fitur kunci dari data mentah. Proses ini menghasilkan representasi yang lebih ringkas dan signifikan, sehingga memungkinkan pemahaman yang lebih baik terhadap data. Dengan melakukan ekstraksi fitur, kita dapat meningkatkan efisiensi dan efektivitas model yang akan digunakan, serta mempersiapkan data untuk analisis yang lebih lanjut.

Fitur yang diambil dapat mencakup atribut-atribut yang krusial atau representasi yang relevan yang membantu dalam proses pengambilan keputusan. Dalam praktiknya, proses ini sering melibatkan berbagai teknik matematis atau statistik, seperti pemilihan fitur yang relevan, transformasi data, dan metode ekstraksi yang lebih rumit. Dengan mengurangi kompleksitas data mentah dan menyoroti informasi yang paling signifikan, *feature extraction* memungkinkan model untuk berfungsi dengan lebih optimal dan memberikan hasil analisis yang lebih akurat, Gambar 2 adalah hasil *feature extractions*.



Gambar 2. Hasil *Feature Extraction*

Dalam eksperimen ini, peneliti memulai dengan inisialisasi matriks `X_train` menggunakan `np.zeros`, yang bertujuan untuk menyimpan fitur-fitur yang diekstraksi dari dataset pelatihan `train_x`. Matriks ini memiliki dimensi `(len(train_x), 3)`, di mana jumlah barisnya sesuai dengan jumlah entri dalam `train_x`, dan jumlah kolomnya adalah tiga, yang menunjukkan bahwa ada tiga fitur yang akan diekstraksi dari setiap entri dataset. Setiap entri dalam `train_x` kemudian diproses secara individual dalam sebuah loop `for`. Untuk setiap entri, data dikonversi ke bentuk numpy array dan diteruskan bersama dengan `freqs` ke dalam fungsi `extract_features`. Fungsi ini bertugas mengekstraksi tiga fitur penting dari data, yang kemudian disimpan dalam baris yang sesuai pada matriks `X_train`. Misalnya, hasil ekstraksi dari entri pertama akan disimpan pada baris pertama dari `X_train`, hasil dari entri kedua pada baris kedua, dan seterusnya.

Setelah semua entri diproses, `X_train` berisi seluruh fitur yang telah diekstraksi dari dataset pelatihan, siap digunakan dalam tahap pelatihan model pembelajaran mesin. Lima baris pertama dari `X_train` ditampilkan untuk memberikan gambaran awal mengenai bentuk dan nilai dari fitur-fitur yang diekstraksi. Dengan demikian, matriks `X_train` menyediakan representasi yang terstruktur dari data mentah, memungkinkan algoritma pembelajaran mesin untuk mempelajari pola dan melakukan prediksi dengan lebih efektif.

3.5. Logistic Regression

Pada tahap pelatihan, model regresi logistik disesuaikan dengan data latih yang telah dilabeli. Tujuannya adalah mengoptimalkan parameter model agar dapat memberikan prediksi yang lebih akurat [24]. Setelah model dilatih, langkah selanjutnya adalah pengujian menggunakan data uji yang belum pernah diakses oleh model sebelumnya. Hal ini bertujuan untuk mengevaluasi kinerja model di luar data latih.

Pengujian akurasi dilakukan untuk membandingkan hasil prediksi model dengan nilai aktual dari data uji. Proses ini memungkinkan peneliti untuk memperoleh gambaran mengenai kemampuan model dalam memprediksi data baru yang tidak dikenal sebelumnya. Penilaian akurasi ini menjadi indikator kunci dalam menentukan efektivitas model [27]. Dengan mengikuti langkah-langkah ini, regresi logistik dapat digunakan untuk menghasilkan model prediktif yang lebih handal dalam berbagai konteks. Implementasi metode ini dapat berkontribusi pada pengembangan analisis yang lebih mendalam dan aplikasi praktis di bidang terkait.

3.5.1. Training

Pada tahap pelatihan model, beberapa parameter penting dihitung untuk menilai kinerja dan hasil dari proses pelatihan. Di bawah ini, disajikan informasi mengenai biaya (cost) setelah pelatihan serta vektor bobot (weights) yang dihasilkan oleh model. Selain itu, terdapat juga peringatan deprecation yang menunjukkan adanya perubahan pada cara konversi array dengan dimensi lebih dari nol menjadi skalar. Tabel 6 merangkum hasil yang diperoleh dari proses pelatihan model.

Tabel 6. Hasil Data Training

Parameter	Nilai
Cost setelah pelatihan	0.2662482
Vektor hasil bobot (weights)	[0.0279653, 0.11597602, -0.24078547]
Peringatan deprecation	Conversion of an array with ndim > 0 to a scalar is deprecated

3.5.2. Testing

Data berikut menggambarkan berbagai entitas dan skor relevansinya berdasarkan analisis tertentu, yang tampaknya terkait dengan pemrosesan teks atau analisis semantik. Setiap baris mengandung daftar kata atau frasa serta skor numerik yang mungkin menunjukkan tingkat keterkaitan atau kepentingannya. Tabel 7 akan membantu menyederhanakan interpretasi data dengan menyajikan setiap komponen secara terorganisir.

Tabel 7. Hasil Data Testing

No	Kata/ Frasa Utama	Skor	Detail Entitas/ Konteks
1	"gimana", "nih", "ceglinya"	0.787175	"mayor", "wkwwk"
2	"evangelista", "pacar"	-0.450044	"mayor", "tedy"
3	"suami", "takluk", "orang"	0.784822	"balas", "benci", "cinta", "rangkuk", "memafkan"
4	"syndrome", "but", "not", "his"	-0.399888	"fault", "so", "you", "cegilecegil", "emakemak", "abdi", "bikin"
5	"mayor"	0.506744	"tedy", "anak", "ngumpul", "prabowo", "sampe", "bingung"

3.5.3 Matrix Evaluation

Berikut adalah hasil evaluasi performa model berdasarkan beberapa metrik utama, yang digunakan untuk mengukur kemampuan model dalam membuat prediksi yang akurat dan relevan. Metrik-metrik ini

mencakup akurasi, presisi, recall, dan F1-score, masing-masing memberikan perspektif berbeda tentang kekuatan dan kelemahan model. Penjelasan mengenai setiap metrik dan nilainya disajikan pada Tabel 8.

Tabel 8. Matrix Evaluation

Metrik Evaluasi	Nilai (%)	Penjelasan
Akurasi	73.68	Persentase prediksi model yang benar terhadap seluruh data uji.
Presisi	75.00	Proporsi prediksi positif yang benar dibandingkan dengan semua prediksi positif.
Recall	71.00	Proporsi data positif sebenarnya yang berhasil teridentifikasi oleh model.
F1-score	73.00	Rata-rata harmonis antara presisi dan recall, mencerminkan keseimbangan keduanya dalam model.

Hasil penelitian menunjukkan bahwa analisis sentimen terhadap Walikota Teddy Indra Wijaya ajudan Prabowo Subianto memberikan gambaran persepsi masyarakat yang terbagi menjadi sentimen positif dan negatif [25]. Secara implisit, analisis ini berperan penting dalam memahami pandangan masyarakat terhadap tokoh masyarakat yang lekat dengan tokoh politik nasional. Sentimen positif tersebut mencerminkan apresiasi masyarakat atas profesionalisme, ketegasan, dan kontribusi Wali Kota Teddy dalam mendampingi Prabowo Subianto. Komentar seperti *“Walikota Teddy terlihat sangat tegas dan kharismatik mendampingi Menteri Prabowo dengan penuh kewibawaan”* dan *“Sosok yang patut dijadikan teladan, Walikota Teddy sungguh inspiratif”* menggambarkan kekaguman terhadap peran Wali Kota Teddy. Selain itu, kehadirannya dinilai mampu mengangkat citra Prabowo Subianto sebagai pemimpin nasional yang berwibawa dan kredibel.

Di sisi lain, sentimen negatif memunculkan isu-isu kontroversial atau berkaitan dengan kehidupan pribadi Wali Kota Teddy. Kritik terhadap sikap yang dianggap kurang profesional, seperti pada komentar *“Sikap Wali Kota Teddy di acara X tidak mencerminkan profesionalisme seorang petinggi ajudan,”* menjadi salah satu sumber utama persepsi negatif. Selain itu, dampak berlebihan yang dirasakan dari promosi diri melalui media sosial juga memicu reaksi skeptis dari sebagian masyarakat. Persepsi negatif ini menyoroti perlunya manajemen yang lebih cerdas dalam menjaga keseimbangan antara keterbukaan di media sosial dan menjaga citra profesional.

Analisis lebih lanjut mengungkapkan bahwa topik yang mendominasi sentimen positif adalah terkait integritas dan kontribusi Wali Kota Teddy, sedangkan sentimen negatif lebih banyak memuat kritik terhadap aspek non-profesional. Dengan mengetahui pola tersebut, strategi komunikasi dapat diarahkan untuk memperkuat aspek positifnya, seperti lebih banyak berbagi cerita inspiratif tentang tindakannya, sekaligus memitigasi dampak isu yang menjadi sumber sentimen negatif. Temuan ini menunjukkan bahwa media sosial dapat menjadi cerminan opini publik yang kuat terhadap figur publik. Hasil analisis ini juga menunjukkan bahwa model Regresi Logistik mampu memetakan sentimen dengan akurasi yang cukup tinggi. Namun, untuk memahami konteks atau makna yang lebih dalam dari komentar yang kompleks, diperlukan pengembangan algoritma yang lebih canggih, seperti metode berbasis *deep learning*.

Secara keseluruhan, analisis sentimen ini memberikan wawasan penting bagi pengelolaan citra publik para tokoh di era digital [26]. Hasil penelitian ini diharapkan dapat menjadi landasan strategis bagi tim komunikasi publik untuk mengoptimalkan citra positif Wali Kota Teddy dan Prabowo Subianto di mata masyarakat. Selain itu, penelitian ini juga membuka peluang penelitian lebih lanjut yang mengintegrasikan analisis tematik guna memperdalam pemahaman persepsi masyarakat yang lebih beragam.

4. KESIMPULAN

Penelitian ini berhasil mengidentifikasi sentimen masyarakat terhadap Walikota Teddy Indra Wijaya, ajudan Prabowo Subianto, dengan menggunakan algoritma Regresi Logistik. Data yang diambil dari platform media sosial Twitter selama Januari 2024 memberikan pandangan mengenai persepsi publik yang terbagi menjadi sentimen positif dan negatif. Hasil analisis menunjukkan bahwa model Regresi Logistik mencapai akurasi sebesar 73,68%, presisi 75%, recall 71%, dan skor F1 0,73. Metrik ini mengindikasikan kemampuan model untuk memprediksi sentimen dengan performa yang cukup baik, meskipun masih terdapat ruang untuk meningkatkan analisis data yang lebih kompleks.

Penelitian ini, bagaimanapun, memiliki beberapa keterbatasan. Pertama, Regresi Logistik kurang optimal dalam menangkap konteks atau makna mendalam dari komentar panjang dan kompleks. Selain itu, jumlah data yang digunakan relatif kecil dan hanya bersumber dari satu platform, yaitu Twitter, sehingga kurang mencerminkan opini masyarakat secara menyeluruh. Keterbatasan ini dapat berdampak pada generalisasi hasil terhadap populasi yang lebih luas atau media sosial lain yang memiliki karakteristik unik.

Untuk penelitian di masa mendatang, disarankan untuk mengadopsi metode analisis yang lebih maju, seperti model berbasis Transformer atau algoritma *deep learning* lainnya. Pendekatan ini dapat memperkaya pemahaman konteks linguistik dan menghasilkan prediksi yang lebih presisi. Memperluas dataset baik dari segi jumlah maupun keragaman platform juga diharapkan mampu menghasilkan temuan yang lebih representatif dan valid. Penggunaan metode kombinasi, seperti *ensemble learning*, dapat menjadi alternatif untuk meningkatkan performa model secara keseluruhan.

Hasil penelitian ini memiliki implikasi strategis dalam merancang komunikasi yang lebih efektif, baik untuk memperkuat sentimen positif maupun mengelola sentimen negatif terhadap figur publik. Pemahaman mendalam melalui analisis sentimen dapat memberikan wawasan penting bagi tim komunikasi dalam mengelola opini masyarakat di media sosial. Dengan memperhatikan keterbatasan yang ada dan mengimplementasikan saran pengembangan, penelitian selanjutnya diharapkan dapat memberikan kontribusi yang lebih besar dalam analisis sentimen pada konteks sosial-politik.

REFERENSI

- [1] K. P. J. Sitompul, A. R. Pratama, and K. A. Baihaqi, "Komparasi Algoritma Naïve Bayes, Support Vector Machine, Dan Logistic Regression Pada Analisis Sentimen Pengguna Aplikasi Transportasi Online," *Kumpulan jurnaL Ilmu Komputer (KLIK)*, vol. 10, no. 1, pp. 27–38, Feb. 2023, doi: dx.doi.org/10.20527/klik.v10i1.616.
- [2] F. Rizal, A. Wijaya, and F. Hasyim, "Analisis Sentimen Masyarakat Indonesia Terhadap Aplikasi TikTok Menggunakan Algoritma Logistic Regression," *AKIRATECH: Journal of Computer and Electrical Engineering*, vol. 1, no. 2, pp. 57–65, Jun. 2024, [Online]. Available: <https://journal.ajbnews.com/index.php/akiratech>
- [3] E. Wibowo and I. Pratama, "Analisis Sentimen Terhadap Ulasan Hotel Melalui Platform Google Review Menggunakan Metode Stacking," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 4, pp. 774–784, Oct. 2024, doi: 10.47233/jteksis.v6i4.1475.
- [4] R. A. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, and L. Efrizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 211–218, Oct. 2023, doi: 10.57152/malcom.v3i2.901.
- [5] S. A. R. Manaf, A. Alamudi, and A. Fitrianto, "Sentiment Analysis of Twitter Users' Opinion Towards Face-to-Face Learning," *Indonesian Journal of Statistics and Its Applications*, vol. 7, no. 1, pp. 15–31, Oct. 2023, doi: 10.29244/ijsa.v7i1p15-31.
- [6] S. Ngeoywijit *et al.*, "Open Innovations for Tourism Logistics Design: A Case Study of a Smart Bus Route Design for the Medical Tourist in the City of Greater Mekong Subregion," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 4, pp. 1–29, Dec. 2022, doi: <https://doi.org/10.3390/joitmc8040173>.
- [7] R. B. Dahlian and D. Sitanggang, "Sentiment Analysis of Digital Television Migration on Twitter Using Naïve Bayes Multinomial Comparison, Support Vector Machines, and Logistic Regression Algorithms," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 280–288, Jul. 2023, doi: 10.32736/sisfokom.v12i2.1668.
- [8] I. R. Ainunnisa and S. Sulastri, "Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan Naive Bayes," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 6, no. 3, pp. 423–430, Jul. 2023, doi: 10.32493/jtsi.v6i3.31076.
- [9] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 154–164, Jul. 2023, doi: 10.37373/tekno.v10i2.440.
- [10] N. P. Agung and H. Bunyamin, "Perbandingan Logistic Regression dengan Random Forest dalam Memprediksi Sentimen Pada IMDb Movie Review," *Jurnal Strategi*, vol. 6, no. 2, pp. 391–399, Nov. 2024.
- [11] A. R. Dewi, S. Diana, M. A. Fakhrezi, N. Awang, H. Ma'arif, and D. D. Saputra, "Sentimen Analisis Terhadap Puan Maharani Sebagai Kandidat Calon Presiden 2024 Berdasarkan Opini Twitter Menggunakan Metode Naive Bayes Dan Adaboost," *JSil (Jurnal Sistem Informasi)*, vol. 10, no. 1, pp. 75–80, Mar. 2023, doi: 10.30656/jsii.v10i1.5785.
- [12] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *JTS*, vol. 1, no. 3, pp. 13–24, Oct. 2022, doi: doi.org/10.56127/jts.v1i3.434.
- [13] A. A. Permana, W. A. Noviyanto, and D. A. Kristiyanti, "Sentimen Analisis Opini Masyarakat Terhadap UMKM Pada Media Sosial Twitter Dengan Metode Naïve Bayes Classifier," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 163–170, Mar. 2023, doi: 10.33395/jmp.v12i1.12337.
- [14] N. A. Amalia, I. T. Utami, and Y. Wilandari, "Analisis Sentimen Kebijakan Penyelenggara Sistem Elektronik Lingkup Privat Menggunakan Penalized Logistic Regression Dan Support Vector Machine," *Jurnal Gaussian*, vol. 12, no. 4, pp. 560–569, Jul. 2024, doi: 10.14710/j.gauss.12.4.560-569.
- [15] N. B. Sidauruk and N. Riza, "Sentimen Analisis Data Pengguna Terhadap Kai Access Systematic Literature Review," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, pp. 1297–1303, Apr. 2023.
- [16] B. Setiawan, K. Ahmad Baihaqi, E. Nurlaelasari, and H. Hikmayanti Handayani, "Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Logistic Regression dan K-

- Nearest Neighbor,” *Technology and Science (BITS)*, vol. 6, no. 1, pp. 533–540, Jun. 2024, doi: 10.47065/bits.v6i1.5389.
- [17] A. Hanafiah *et al.*, “Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier Sentiment Analysis Of Customer Reviews Of Shopee Products Based On Wordcloud Using Naïve Bayes Classifier Algorithm,” *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, pp. 230–236, Jun. 2023, doi: 10.31539/intecom.v6i1.5845.
- [18] F. Maulana Herza, B. Rahmat, and M. Muharrom AL Haromainy, “Pengaruh RFE Terhadap Logistic Regression Dan Support Vector Machine Pada Analisis Sentimen Hotel Shangri-La Surabaya,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 6, pp. 11612–11619, Oct. 2024, doi: doi.org/10.36040/jati.v8i6.11272.
- [19] Alisya Mutia Mantika, Agung Triayudi, and Rima Tamara Aldisa, “Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election,” *SaNa: Journal of Blockchain, NFTs and Metaverse Technology*, vol. 2, no. 1, pp. 44–55, Feb. 2024, doi: 10.58905/sana.v2i1.267.
- [20] E. G. Radjah and A. C. Talakua, “Analisis Sentimen Komentar Terhadap Konten Tenun NTT di Youtube Menggunakan Metode SMOTE dan Logistic Regression,” *Jurnal Transformatif*, vol. XIII, no. 2, pp. 84–94, Oct. 2024, doi: doi.org/10.58300/transformatif.v13i2.
- [21] A. Hagi and D. B. Rarasati, “Sentiment Analysis of Sirekap Application Review Using Logistic Regression Algorithm,” *Jurnal Informatika*, vol. 11, no. 2, pp. 55–64, Aug. 2024, doi: 10.31294/inf.v11i2.22066.
- [22] N. Agustina and C. N. Ihsan, “Pendekatan Ensemble Untuk Analisis Sentimen Covid19 Menggunakan Pengklasifikasi Soft Voting An Ensemble Approach For Covid19 Sentiment Analysis Using Soft Voting Classifier,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 2, pp. 263–270, Apr. 2023, doi: 10.25126/jtiik.2023106215.
- [23] M. Putri Agustina, “Sentimen Masyarakat Terkait Perpindahan Ibukota Via Model Random Forest dan Logistic Regression,” *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 111–124, Nov. 2021, doi: doi.org/10.24246/aiti.v18i2.111-124.
- [24] A. Tyagi and N. Sharma, “Sentiment analysis using logistic regression and effective word score heuristic. International Journal of Engineering and Technology (UAE), 7(2.24), 20–23.,” *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 2, pp. 20–23, 2018.
- [25] J. S. Hutagalung and Rasiban, “Analisis Sentimen Keuangan (Data Fiqa And Financial Phrasebank) Menggunakan Algoritma Logistic Regression Dan Support Vector Machine,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1654–1669, Sep. 2023, doi: 10.35870/jimik.v4i3.404.
- [26] I. A. Mashudi and S. N. Arief, “Analisis Sentimen Perkembangan Kasus Covid-19 Pada Komentar Facebook,” *Jurnal Teknik: Ilmu dan Aplikasi*, vol. 02, no. 1, pp. 5–9, Jan. 2021, [Online]. Available: www.abc.com
- [27] W. J. Sari *et al.*, “Performance Comparison of Random Forest, Support Vector Machine and Neural Network in Health Classification of Stroke Patients,” *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 1, pp. 34–43, Apr. 2024, doi: 10.57152/predatecs.v2i1.1119.